

Listing of Claims:

1. (Currently amended) A method in a data processing system for generating and storing in a database an entry, the method comprising the steps of:

generating an entry comprising:

i) data identifying a molecule;

ii) data identifying at least one region in the molecule; and

iii) a set of axes derived from property distribution information of the at least one region, the set of axes characterizing the at least one region;

generating at least one descriptor vector for the at least one region;

applying a mapping to the at least one descriptor vector associated with the at least one region to construct a key based on preselected criteria; and

storing the entry in a memory, wherein the key is associated with the entry such that the key indexes the entry for retrieval thereof.

~~— In a data processing system wherein descriptor vectors associated with a plurality of regions of molecules are stored in a database, a method for generating and storing data characterizing at least one region of said plurality of regions, the method comprising the steps of:~~

~~— generating an entry comprising i) an identifier that identifies said at least one region, and ii) data characterizing a set of axes derived from a property distribution of said at least one region;~~

~~— applying a mapping to the descriptor vector associated with said at least one region based on preselected criteria;~~

~~— generating a key that corresponds to said mapping of the descriptor vector associated with said at least one region; and~~

~~— storing said entry in a memory, wherein said key is associated with said entry such that the key indexes the entry for retrieval thereof.~~

2. (Cancelled)

3. (Cancelled)

4. (Currently amended) The method of claim 1, wherein said the property distribution information of said the at least one region is computed from a convolution with a probe function to a property field.

5. (Currently amended) The method of claim 1, wherein said the at least one plurality of descriptor vectors are is classified into groups, and wherein said the mapping step maps said the at least one descriptor vectors to a space discriminating between said groups of descriptor vectors.

6. (Currently amended) The method of claim 5, wherein said the mapping is derived from the steps of:

generating first data representing differences between said groups of descriptor vectors;
generating second data representing variations within said groups of descriptor vectors;

identifying a set of component vectors that maximizes a ratio of variations between groups to the variations within groups along the component vectors as a discriminant criterion function ~~an F-distributed criterion function, said criterion function having a numerator based upon said first data and a denominator based upon said second data;~~

generating a criterion function for subsets of the component vectors, wherein the criterion function utilizes the first data and the second data ~~an F-distributed statistic for subsets of said component vectors, said statistic having a numerator based upon said first data and a denominator based upon said second data;~~

for each particular subset of ~~said~~ component vectors, calculating a probability value for the ~~F-distributed statistic~~ criterion functions associated with the particular subset;

selecting a probability value from probability values for said the subsets of said component vectors based upon a predetermined criterion;

identifying the subset of said-component vectors associated with the selected probability value; and

generating a mapping to a space corresponding to the subset of said-component vectors associated with the selected probability value, and storing the mapping for subsequent processing.

7. (Currently amended) The method of claim 6, wherein said the first data comprises a matrix ϵ_b representing covariance between said the groups of descriptor vectors, and said the second data comprises a matrix ϵ_w representing covariance within said the groups of descriptor vectors.

8. (Currently amended) The method of claim 7, wherein said the criterion function has the general form:

$$f(\hat{w}) = C \left(\frac{\hat{w}^T \epsilon_b \hat{w}}{\hat{w}^T \epsilon_w \hat{w}} \right)$$

where \hat{w} is some vector, T indicates a transpose, ϵ_b is a first data representing covariance, ϵ_w is a second data representing covariance and C is a constant based upon degrees of freedom in ϵ_b and ϵ_w .

9. (Currently amended) The method of claim 8, wherein the variable C is determined as follows:

$$C = \frac{1/\text{degrees of freedom in } \epsilon_b}{1/\text{degrees of freedom in } \epsilon_w} = \frac{1/(N-1)}{1/(\sum n_i - N)}$$

where N represents the number of groups of descriptor vectors, n_i represents the number of regions, and $\sum n_i$ represents the sum of n_i for the N groups.

10. (Currently amended) The method of claim 7, wherein the step of identifying a set of component vectors that maximizes an ~~F-distributed~~ F-distributed criterion function comprises the substeps of:

determining a set of (eigenvalue, eigenvector) pairs for the matrix ϵ_w ; and

determining ~~said the~~ set of component vectors based upon ~~said the~~ set of (eigenvalue, eigenvector) pairs for the matrix ϵ_w .

11. (Currently amended) The method of claim 10, wherein ~~said-statistic the F-~~ distributed statistic for a given subset of component vectors is based upon value of ~~said the~~ criterion function for said the subset of component vectors.

12. (Currently amended) The method of claim 11, wherein ~~said-statistic the F-~~ distributed statistic for a given subset of component vectors has the following form:

$$\Psi_S = C \left(\frac{1}{L_S} \right) \sum f_k$$

where f_k represents the value of the criterion function at a component vector in the given subset, C is a constant, L_S represents the number of f_k values in the given subset of component vectors, and the \sum operation sums over the L_S f_k values in the given subset of component vectors.

13. (Currently amended) The method of claim 12, wherein ~~said a~~ the probability value for a particular F-distributed statistic represents a probability value that the particular F-distributed statistic could have been larger by chance.

14. (Currently amended) The method of claim 13, wherein ~~said~~ the probability value selected from probability values for ~~said the~~ subsets of component vectors is a minimum probability value of ~~said the~~ probability values for ~~said the~~ subsets of component vectors.

15. (Currently amended) The method of claim 6, wherein ~~said~~ the mapping for ~~said the~~ at least one descriptor vector performs a loop over each component vector belonging to the subset of component vectors associated with the selected probability;

wherein, in each iteration of ~~said the~~ loop, dot product of ~~said the~~ descriptor vector with a transpose of a unit vector for the given component vector is added to a running sum.

16. (Cancelled)

17. (Cancelled)

18. (Cancelled)

19. (Cancelled)

20. (Cancelled)

21. (Cancelled)

22. (Cancelled)

23. (Cancelled)

24. (Cancelled)

25. (Cancelled)

26. (Cancelled)

27. (Cancelled)

28. (Cancelled)

29. (Cancelled)

30. (Cancelled)

31. (New) The method of claim 1, wherein the at least one descriptor vector is invariant to rotation and translation of the at least one region.

32. (New) The method of claim 31, wherein the set of axes is derived from principal axes of second moments of a region of the property distribution information.

33. (New) The method of claim 6, wherein the probability value is obtained by treating the ratio as an F-distributed statistic.

34. (New) The method of claim 6, wherein the probability value is obtained by any one of cross-validation, jack-knife and bootstrap estimations.

35. (New) The method of claim 6, wherein application in constructing the discriminant criterion function includes boosting and bagging techniques.